

ソーシャルシェアデータを用いた観光エリア推薦システム

Sightseeing-Area Recommendation Based on Social Data

加藤風太^{1*} 熊野雅仁² 木村昌弘²
Futa Kato¹ Masahito Kumano² Masahiro Kimura²

¹ 龍谷大学大学院理工学研究科電子情報学専攻

¹ Division of Electronics and Informatics, Ryukoku University

² 龍谷大学理工学部電子情報学科

² Department of Electronics and Informatics, Ryukoku University

Abstract: 近年、Web空間に共有化されたGeo-tag付ビッグデータを観光に応用する研究が注目されている。従来研究では、主に観光スポットが着目されてきた。我々は、ユーザの実世界における行動履歴情報を集合知的観点から集約することで、個人化推薦システムの構築を目指している。本研究では、訪れる地域が指定されたとき、大量のGeo-tag付き写真に基づいて複数の観光スポットを含む地域を効果的に可視化・推薦する観光エリア推薦システムを提案する。

1 はじめに

近年、ソーシャルメディアの発展により、ユーザが日常行動に応じて感じた思いや気持ちを言語化したり、評点、賛意の表明など、様々なアクティビティを通じてWeb空間に公開し、シェアする時代が到来しており、人々の多様な意見や嗜好情報を含んだソーシャルシェアデータがビッグデータとして注目されている。

一方、人々の観光行動においても、観光先で、実際に、何にどのような魅力を感じたかについての情報がWeb空間で公開され、シェアされ始めているため、エビデンス（実行動）に裏づけられた、人々の観光に関する潜在的な嗜好を抽出し、有益な情報を旅行者に還元できる新たなサービス創出の可能性が高まっている。

これまで、観光産業では、独自の調査と専門的な知識に基づいて観光案内情報を収集・集約し、魅力的な観光先の推薦を行ってきた。しかし、人々の多様で潜在的な嗜好を捉えることができれば、これまでの観光産業による観光先や、多くの人に好まれる観光先に加え、観光の専門家が気づけなかった、もしくは注目しなかった観光先を嗜好に応じて適切に推薦できる可能性があるため、個人化され、より洗練された推薦システムを構築できる可能性が期待される。また、これまで、観光対象を検索するシステムでは、主に個々の観光スポットが独立に扱われるため、例えば、検索結果として得られた観光スポット上位二つが、実空間上、距離的に離れているため、ユーザの持ち時間では二つを

訪れることができない問題などが生じる場合も考えられる。つまり、観光先に関する推薦では、個々の観光スポットを推薦するよりも、観光スポットが複数含まれる、実空間上のエリアを推薦することが望ましいと思われる。さらに、例えば「お寺好き」「街歩き好き」「お酒好き」など、ユーザは一般に多重嗜好を持つので、ユーザの多重嗜好を効率良く反映できる推薦システムが望まれる。

本研究では、ユーザが訪れる都市を指定したとき、ソーシャルメディアデータとしての何年にも渡る大量のGeo-tag付き写真データから集合知的観点に基づいて、多くの人々に好まれる観光先だけでなく、行動パターンが近い撮影者達の訪問先も強調して複数同時に可視化することで、ユーザの多重嗜好に応じた観光スポットを地図システム上に可視化する観光エリア推薦システムを提案する。2010年から2014年までの日本で撮影されたGeo-tag付写真による実データを用いて、評価実験を行い、提案する観光エリア推薦システムの有効性を示す。

2 観光エリアの個人化推薦

個人の嗜好に応じた適切な推薦を行う上で、推薦システムを使うユーザの嗜好データがない場合や、推薦する対象に嗜好データがない場合、ユーザへの推薦がうまくいかないというコールドスタート問題がある[1]。コールドスタート問題の解決は重要な課題となっている。

また、例えば、訪れる都市が決まっているものの、その都市や周辺地域に存在する施設や観光先を知らない場合を考える。予め、ユーザに関する嗜好情報が得られ

*連絡先：龍谷大学
滋賀県大津市 瀬田大江町横谷 1-5
E-mail:t14m008@mail.ryukoku.ac.jp

ていない場合、ユーザ自身の好みが見えやすい場合は、ユーザに、好みを限定するためのカテゴリ選択やキーワードの入力を要求する方法が考えられる。しかし、現地に何が存在するか知らない場合や、ユーザ自身が自分の嗜好をうまく把握できていない場合など、カテゴリの選択が難しく、検索用のキーワードが思い浮かばない場合さえある。したがって、多くのユーザにとっては見やすいキーワードやカテゴリを特定して観光先を絞ることが一般に困難である。

一方、近年、Trip Advisor や Yelp, Foursquare など、施設への評価を投稿する施設共有サイトが注目されている。ユーザがこれらのサイトにおいて、過去に訪れた施設への評価を登録していれば、嗜好情報として推薦システムに与える方法を考えることができる。ただし、例えば、近年の訪日外国人が興味を示す観光スポットとして、東京・渋谷のスクランブル交差点や、新宿ゴールデン街の街並みなど、見やすい施設ではない意外な場所が観光先となっていたり、京都・伏見稲荷大社の千本鳥居など、大きな敷地に存在する施設の一部に人気が集まる場合もあるため、きめ細かな嗜好を捉えるには、施設という単位に依存しない柔軟な観光先の捉え方と、その観光先に関する嗜好を捉え得る手法が望まれる。

我々は、人々が多重の嗜好を持つ傾向があると考えており、個人向けのきめ細かい推薦を行うためには、指定された都市や周辺地域について、見やすいキーワードやカテゴリを特定して観光先を絞り単一の観光スポットを探し当てるのではなく、嗜好の近さを可視化しつつ多様な観光スポットを同時に提示することが望ましいと考えている。

これらの観点に対して、本研究では、人々の写真撮影行動に着目する。近年、GPS に基づく情報 (Geo-tag) を写真に付与できるカメラやスマートフォンの普及とともに、Flickr などの写真共有サイトが普及することで、一般の旅行者達が観光した際に撮影した大量の写真が、撮影場所の緯度経度、撮影日時、焦点距離などの撮影条件、付与されたコメント、撮影者のプロフィールやソーシャルネットワーク情報などとともに、Web 空間に蓄積され続けており、世界的に大規模なソーシャルシェアデータとなっている。旅行者は、視覚的に興味を抱いた対象に出くわすと写真を撮影する傾向があり、厳選した写真を Web 空間に公開する傾向があると考えられるため、質の高い嗜好情報が得られる可能性がある。そのため、写真の撮影行動を集約し、人気スポットを抽出する研究 [2] や、観光へ応用する研究が注目されている [3], [4], [5], [6], [7]。つまり、Geo-tag を用いた集合論的観点から観光スポットを抽出するアプローチを採用し、嗜好が似た撮影者群を捉えることで、施設名を持たない意外な観光スポットも抽出できる可能性が期待される。また、ユーザのカメラやスマートフォ

ンに蓄積された過去の写真群や、写真共有サイトに公開した写真と付随する情報をユーザの嗜好データとして推薦システムに与えれば、撮影者が自らの嗜好をうまく言語化できない場合でも、嗜好に応じた観光先を推薦することができる可能性がある。さらに、ユーザの嗜好に合う多様な観光スポットを嗜好の近さが判別できるように地図上に同時に可視化すれば、観光において、現実的に使用できる制限された時間帯で、どのエリアを観光するかについて、効率的な選定を可能にするシステムの実現が期待される。本研究では、これらの観点に基づいた観光エリア推薦システムを提案する。

3 提案システム

3.1 システム概要

ユーザが訪れる都市の観光エリアを推薦する問題に対して、我々は、まず、明確な領域を持つ施設単位で観光スポットを捉えるのではなく、過去において、人々の実際の行動に裏づけられた未知数の観光スポットを含む観光エリアを自動的に抽出するため、大量の Geo-tag 付写真データから観光エリアを抽出する。ここで、既存の施設は土地画面上の領域を持つが、提案法での観光エリアは実行動に基づいて定まる領域であることに注意しておく。抽出された多数の観光エリアのうち、撮影者の人数が多いほど、人気エリアであると見なせる。人気エリアは、どのような嗜好のユーザに対しても、基本的な推薦対象と考えられる。また、全く嗜好情報が得られないユーザに対して人気エリアを推薦すれば、ユーザに関するコールドスタート状態でも、推薦先が無い状況を避けられる点に注意しておく。

次に、過去の撮影行動情報が得られるユーザに対し、観光エリアの個人化推薦問題におけるユーザの多重嗜好に配慮した推薦手法の第一歩として、本研究では、協調フィルタリングの観点に着目する。例えば、京都を訪れる予定のユーザ u が過去に北海道で写真を撮影した観光エリアにおいて、同様に撮影を行った他のユーザ w が既に京都を訪れ写真を撮影した観光エリアが存在した場合を考える。このとき、協調フィルタリングでは、ユーザ u とユーザ w の過去の撮影行動が似ているほど、類似度が高い関係と見なし、ユーザ w が過去に訪れた観光エリアを推薦する。ここで、ユーザ u が多様な撮影行動をしている場合、ユーザ u の異なる嗜好ごとに類似性の高いユーザ w, w' が、過去に京都を訪れていれば、ユーザ u が訪れる京都で、多様な嗜好に基づくエリアを類似性が高い観光エリアとして同時に推薦できる可能性がある。

本研究では、人気エリアと多重嗜好を考慮したエリアの両方を重視することから、両者を統合して観光エリアとしてユーザに提示する推薦手法を提案する。

3.2 観光エリア抽出

施設単位の観光スポットではなく、過去の撮影行動に基づいて観光エリアを抽出するため、ここでは Mean-Shift 法 [8] に基づく観光エリア抽出法を適用する。また、ユーザが指定した都市に対し、観光エリアを推薦するための、各種推薦手法と提案推薦法について述べる。

3.2.1 入力データ

正の整数 T に対して、 T 日の期間 $[1, T]$ 内に撮影された写真データ全体の集合を \mathcal{D}_0 とする。本研究では、集合知的観点から個人差を吸収し、撮影者の人数を重視するため、緯度と経度 2 の 2 次元平面を離散化し、最小矩形領域ごとに、1 人の撮影者による写真を 1 枚抽出する正規化を行う。このとき、最小矩形領域内の写真数は、その領域内で撮影を行った人数に相当することを注意しておく。この離散正規化を \mathcal{D}_0 に適用して得られる写真集合を、

$$\mathcal{D}_1 = \{d_n; n = 1, \dots, N\}$$

とする。また、 \mathcal{D}_1 の写真を撮影したユーザ集合を

$$U_1 = \{u_i; i = 1, \dots, M\}$$

とする。各写真データ d_n には位置情報 x_n 、時間情報 t_n 、ユーザ情報 u_m とが付随しており、

$$d_n = (x_n, t_n, u_i), (n = 1, \dots, N, i = 1, \dots, M)$$

と記述する。ただし、 $x_n = (x_{n,1}, x_{n,2})$ であり、 $x_{n,1}$ と $x_{n,2}$ はそれぞれ緯度と経度、 t_n は d_n が撮影された日、 N 写真データ総数、 M はユーザ総数である。

3.2.2 Meanshift 法に基づく観光エリア抽出法

d 次元 Euclid 空間 \mathbf{R}^d 上の点群 $S = \{s_n\}_{n=1}^N$ がある確率分布に従う標本集合であるとき、任意の点 $s \in \mathbf{R}^d$ における確率密度関数を、ノンパラメトリックアプローチであるカーネル密度推定

$$\hat{p}(s) = \frac{1}{N} \sum_{n=1}^N \frac{1}{h^d} G\left(\left\| \frac{s - s_n}{h} \right\|^2\right), (s \in \mathbf{R}^d)$$

により推定することを考える。ここに、 $\|\cdot\|$ は \mathbf{R}^d の Euclid ノルム、 $G(z)$ はカーネル関数である。ここで、 $G(z)$ は、Epanechnikov カーネルを利用する。また、 $h (> 0)$ は、データの各位置 s_n ごとに存在する確率密度関数のバンド幅を規定するパラメータである。Crandall ら [2] は、写真データ集合 \mathcal{D}_1 から主要な撮影地域を抽出する手法として、緯度と経度に基づく $d=2$ 次元の点

群 $\{x_n\}$ を対象として MeanShift クラスタリングを適用している。

ところで、主要な撮影地域は、街角レベルから都道府県、州、国規模など、様々なスケールが考えられるため、最適なサイズを容易に決定できない問題がある。本研究では、徒歩圏内の多くの撮影者が集まる地域を観光エリア A_k と呼び、Crandall らの Metropolitan-scale ($h=100\text{m}$) として Epanechnikov カーネルを用いた Meanshift 法 [8] によって観光エリア A_k の抽出を行う。ここで、観光エリア A_k は、その領域に含まれる写真群の撮影位置 $x_n \in A_k$ の集合とするが、撮影位置の集合に基づく地理空間上の領域を示しているとする。また、 Δt_0 年内 (期間 I_0 と呼ぶ) に撮影された写真群を $\mathcal{D}_2 = \{d_n \in \mathcal{D}_1; t_n \in I_0\}$ とするとき、 \mathcal{D}_2 を対象に Meanshift 法で抽出された観光エリア集合を $\mathcal{A}_0 = \{A_k; k = 1, \dots, K'\}$ とする。また、 A_k に含まれる写真集合を $\mathcal{D}_k = \{d_n \in \mathcal{D}_2; x_n \in A_k\}$ としたとき、 $|\mathcal{D}_k| > \mu_0$ を満たす観光エリアを

$$\mathcal{A}_1 = \{A_k; k = 1, \dots, K\}$$

と記述する。ただし、 K は抽出された観光エリアの総数である。

3.2.3 新規撮影地点の観光エリア配属法

ユーザ u_i が撮影した観光エリアと、他の撮影者 u_j ($i \neq j$) が撮影した観光エリア \mathcal{A}_1 が一致するかを調べる方法を述べる。本研究では、提案法を評価する際、ユーザ u_i が、新たな期間 I_1 ($I_0 \cap I_1 = \emptyset$) に訪問する観光エリア A_k を予測することで、推薦システムの性能評価を行う。ただし、期間 I_0 で抽出される観光エリアと、期間 I_1 で抽出される観光エリアは同じ Meansift 法を適用しても、全く同じになるとは考えにくい。期間 I_0 には全く撮影が行われていなかったエリアが I_1 で撮影が行われるようになったり、ほぼ同じエリアでも、期間の違いにより、写真数や撮影位置が変化し得るため、完全に領域が一致しないものがほとんどであると予想される。このとき、ユーザ u_i が、期間 I_1 に撮影を行った位置と、他のユーザ u_j が撮影を行った位置が同じ撮影エリアであるかを定める上で、曖昧性が生じる。そこで、本研究では、期間 I_1 に撮影された写真がどの観光エリアに帰属するかは、過去の期間 I_0 で抽出された観光エリア \mathcal{A}_1 に帰属するかを調べるという方法を採用する。これは、期間 I_0 のデータで生成した確率密度関数を用いて新規撮影地点の収束先を求め、確率密度関数の極値近傍に収束したか否かを判別する方法となる。つまり、いずれかの極値の近傍と判断すれば新規撮影地点の配属先が定まる。ただし、いずれの極値近傍でもない判断されれば、配属先がないことになるが、これは、新しく表れた観光エリアである可

能性があるものの、本研究における予測実験においては対象外と見なす。

3.3 観光エリア推薦法

ユーザ u_i から訪れる都市 c が指定されたとき、本研究では、 A_1 のうち、 c に含まれる観光エリア集合を A_1^c 、 c 以外の都市 c' ($c \neq c'$) に含まれる観光エリア集合を $A_1^{c'}$ とする。このとき、各観光エリア $A_k \in A_1^c$ に優先順位を与え、ランキング形式で観光エリアを推薦する手法を採用する。ここでは、優先順位を与える方法として、人気エリア法、二つの協調フィルタリング法、混合法について述べる。

3.3.1 人気エリア法

抽出された観光エリア A_k に含まれる写真群の数 $|D_k|$ が多いほど人気度の高い観光エリアであると言える。つまり、推薦する観光エリア A_k のスコア AS_k を

$$AS_k = |D_k|$$

とし、人気度の高さに応じて AS_k をランキングする手法である。この手法では、たとえユーザ u_i に関して、過去の撮影行動データが無いコールドスタート問題が生じる場合でも、多くの人々に人気のある観光エリアを推薦することができる。ただし、すべてのユーザに対して同じ推薦結果となることに注意しておく。

3.3.2 協調フィルタリング法

ユーザ u_α へ観光エリアを推薦する際、ユーザ u_α と行動類似性の高い、他のユーザ u_β が過去に撮影した観光エリアを推薦する手法である。ユーザ u_α と、ユーザ u_α 以外のユーザ u_β ($\alpha \neq \beta$) の行動類似性を $sim(u_\alpha, u_\beta)$ で表す。ユーザ u_β が都市 c でこれまでに撮影したことがある観光エリアの全体の集合を $RA_{u_\beta}^c$ とする。このとき、推薦する観光エリア A_k のスコア AS_k を

$$AS_k = \sum_{u_\beta \in U_1} \chi_{k,\beta} sim(u_\alpha, u_\beta)$$

で定義する。ここに

$$\chi_{k,\beta} = \begin{cases} 1 & (A_k \in RA_{u_\beta}^c) \\ 0 & (A_k \notin RA_{u_\beta}^c) \end{cases}$$

とする。これは、投票形式で観光エリア $A_k \in RA_{u_\beta}^c$ にスコアを加える方法となるが、ユーザ u_α と行動類似性の高い他のユーザが撮影した観光エリアほど、高い値が加算される点で嗜好が考慮されていくことにな



図 1: 観光エリア A_k のスコア AS_k と色の対応

る。つまり、スコア AS_k の大きさに基づいて観光エリアのランキングを行い、推薦を行う手法が本研究における協調フィルタリング法である。また、ユーザ u_β ごとに、異なる嗜好に基づいた類似度がスコア AS_k へ加算されることから、多重嗜好を反映した複数の観光エリアを推薦できる可能性があることに注意しておく。

jaccard 係数

ユーザ u_α が期間 I_0 で撮影経験のある観光エリアの集合を $PA(u_\alpha)$ 、ユーザ u_α 以外のユーザ u_β ($\alpha \neq \beta$) が撮影経験のある観光エリアの集合を $PA(u_\beta)$ とするとき、 $PA(u_\alpha)$ と $PA(u_\beta)$ の共起性の度合いを表す jaccard 係数を用いて $sim(u_\alpha, u_\beta)$ を表す方法である。

$$sim(u_\alpha, u_\beta) = \frac{|PA(u_\alpha) \cap PA(u_\beta)|}{|PA(u_\alpha) \cup PA(u_\beta)|} \quad (1)$$

3.3.3 混合法

本研究では、ユーザの過去の行動履歴が無い場合や、多くの人に好まれる観光エリアが推薦できる人気エリア法に加え、ユーザの過去の行動履歴がある場合には、個人の嗜好に寄り添った推薦を行う方法を実現するため、人気エリア法と協調フィルタリング法の混合法を提案する。 A_k に対し、正規化された人気エリア法のスコアを ξ_k 、正規化された協調フィルタリング法のスコアを ϕ_k とし、 $0 \leq w \leq 1$ としたとき、次のようにする。

$$AS_k = (w - 1)\xi + w\phi$$

3.4 観光エリアの可視化法

提案する観光エリアの可視化法としては、ユーザ u_α から指定された都市 c に対し、都市 c に含まれる抽出された観光エリア A_k を可視化する。ただし、観光エリア A_k は、撮影地点の集合であり、抽出された A_k は領域を持っている。そこで、提案法では、観光エリアを A_k の代表地点で表現するのではなく、撮影地点を包含する円領域として可視化することにより観光エリアが地図上において占める領域の相対的な大きさを容易に視認できる可視化法を採用する。円領域の決定法としては、観光エリア A_k 内の撮影地点群 D_k の中心点を定め、その中心から最も遠い撮影位置を半径とした円の内側を観光エリア A_k の領域として可視化に用いる。

また、 A_k の領域と同時に、ユーザ u_α の嗜好に合う観光エリア A_k を強制的に可視化し、特徴的なお勧め観

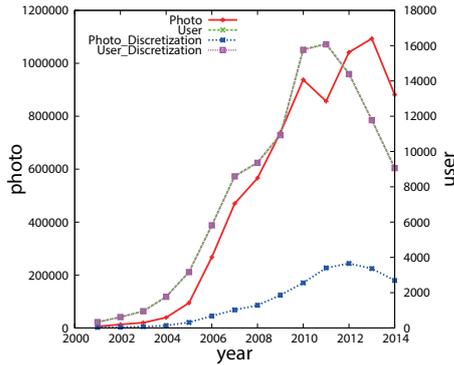


図 2: 写真共有サイト Flickr に登録された日本における Geo-tag 付写真数とユーザ数の変遷

表 1: 初期データセット

DataSet	DataSet'	訓練&学習データ	予測データ (京都)
Dataset1	Dataset1'	2010 年, 2011 年	2012 年
Dataset2	Dataset2'	2011 年, 2012 年	2013 年
Dataset3	Dataset3'	2012 年, 2013 年	2014 年

観光エリアの発見を促す可視化システムを提案する。観光エリア A_k を嗜好に応じて強制的に可視化する方法として、ユーザ u_α の嗜好との類似性がスコア AS_k として算出されているため、本研究では、図 1 に示した色の対応を用い、スコア AS_k が低いほど青く高いほど観光エリア A_k の領域を赤く提示することで観光エリア推薦度を強制的に可視化する。これにより、都市 c において、一般的な観光エリアとともに、ユーザ u_α の嗜好に合う複数の観光エリア候補から、現実的な持ち時間や移動距離を考慮して訪問先を吟味することができる。

4 エリア抽出法及び推薦法評価実験

4.1 データセット

日本を対象に提案システムの検証実験を行うため、写真共有サイト Flickr から、日本の WoID (23424859) を持ち、日本国内で位置情報を持ったデータの収集した。図 2 は収集した写真データについて、2000 年から 2014 年までの位置情報付写真数とユーザ数の変遷である。また、離散化を行った結果の写真数とユーザ数の変遷も同様に図 2 に示す。図 2 より 2010 年から 2014 年において写真数が増加しているが、本論文では、投稿が盛んな期間を対象として実験データセットの構築を行う。また、ユーザが指定する都市としては、数多くの多様な観光エリアが存在する京都を対象として実験を行う。実験では 2 年間で訓練データとして、翌年

表 2: 観光エリア抽出後の最終データセット

Dataset	観光エリア	訓練ユーザ	予測ユーザ数
Dataset1	2548	850	620
Dataset2	3122	1050	620
Dataset3	3440	1399	637
Dataset1'	2548	850	560
Dataset2'	3122	1050	582
Dataset3'	3440	1399	604

表 3: 各データセットの観光エリア $|A_k|$ 抽出数と配属率

Dataset	観光エリア $ A_k $	配属率 (%)
Dataset1	108547	69.0
Dataset2	117329	70.1
Dataset3	107779	67.9
Dataset1'	108547	74.0
Dataset2'	117329	76.1
Dataset3'	107779	74.5

に京都を訪れているユーザを対象に実験を行うためのデータセットを構築した。また、学習データにおいてユーザが京都を訪れている、つまり過去に京都の観光エリアを訪れている場合に推薦において影響がある可能性を考慮し、その情報を削除した Dataset' を構築した。データセットの詳細を表 1 に示す。ただし、抽出された観光エリアを用いて観光エリアの推薦を行う上で、本研究では観光エリアが多数の嗜好を反映している必要を考え、 $\mu_0 \geq 10$ を満たす観光エリアを A_k とした。また、協調フィルタリングを行う上で推薦を行うためには訓練ユーザが予測対象都市 c とその他に少なくとも一つの観光エリアを訪れている必要がある。以上の観点から最終的なデータセットは表 2 のようになった。

4.2 観光エリア抽出結果と配属率の結果

ここで、表 1 の各データセットについて観光エリア A_k の抽出を行った結果を表 3 に示す。また、ユーザ $u \in U_1$ を対象として、予測期間にユーザ u が撮影した地点が抽出されたいずれかの観光エリア A_k に配属される配属率を求めた。各データセットに関する配属率の結果も表 3 に示す。どのデータセットにおいても 7 割前後と比較的高い水準で配属されていることがわかる。

4.3 評価手法

本研究では、予測期間に京都を訪れているユーザが実際に撮影した観光エリアを隠蔽し、ユーザへ推薦する観光先をスコア AS_k のランキング上位から順に推薦したとき、ユーザが実際に訪れた観光エリアを正解データとして、推薦先と一致するかという観点から、適合率 (precision) に着目して評価を行った。また、提案

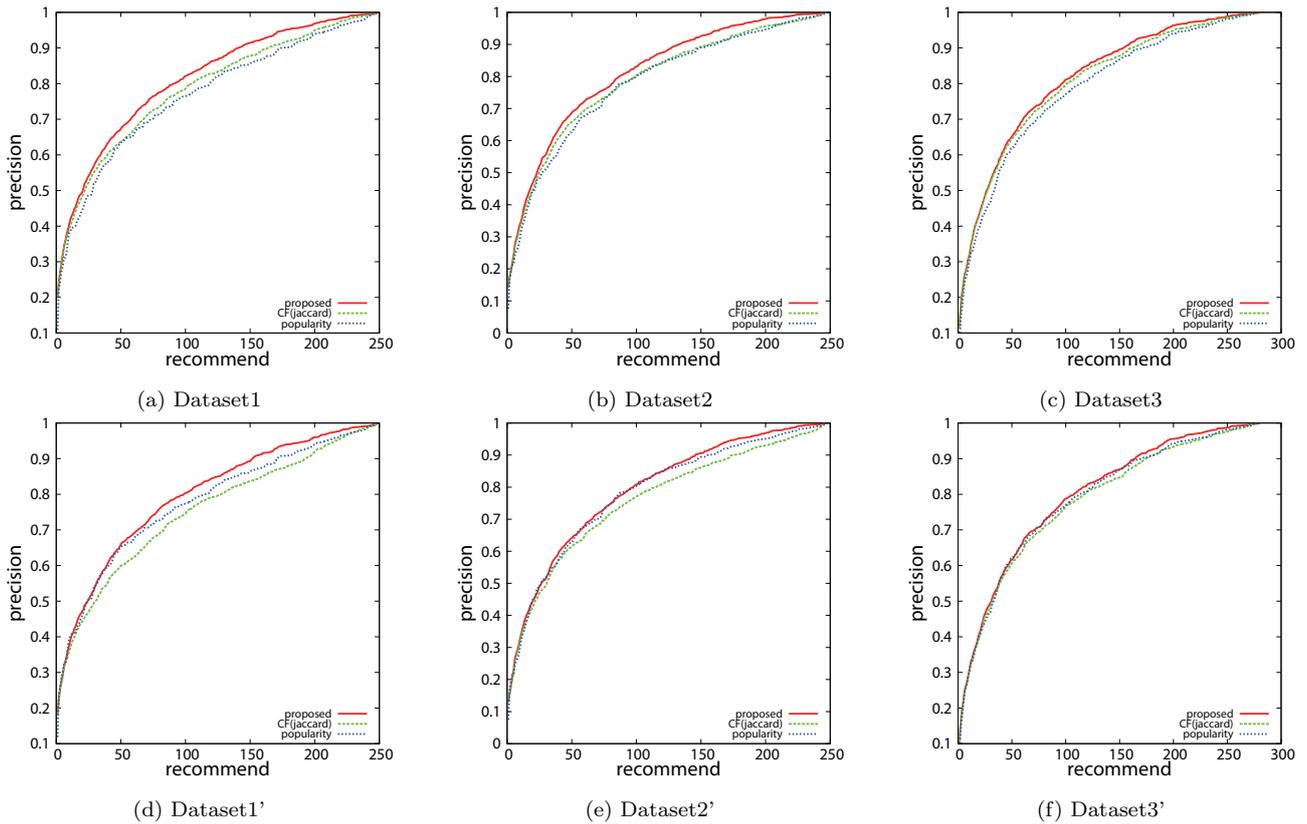


図 3: 各 Dataset における観光エリア推薦法適用結果の precision

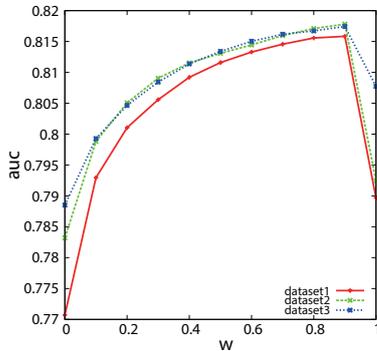


図 4: Dataset におけるパラメータ w と AUC 値の関係

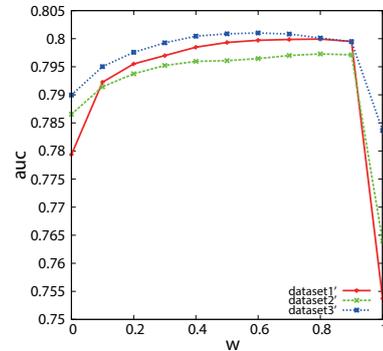


図 5: Dataset' におけるパラメータ w と AUC 値の関係

する混合法において、AUC 値が最も高い値を示す重み w を検証した。

4.4 実験結果

図 3 は各データセットに対し評価実験を行った際の precision の結果である。図 3 において、どのデータセットにおいても協調フィルタリングと人気エリア法が共に高い結果を示している。特に図 3(a), (b), (c) においては協調フィルタリングを用いた値が高く、個人化推薦の有用性が示唆されるが、図 3(d), (e), (f) においては人気エリア法に比べ協調フィルタリングがわず

かに低い値を示している。次に提案混合法を用い、協調フィルタリング法 (jaccard 係数) に人気エリア法を混合することによりアクティビティの少ないユーザへの推薦を補完し、コールドスタート問題に対応した推薦を行う。そのために混合法における最適な重み w を決定すべく実験を行った。その結果を図 4 と図 5 に示す。実験結果、図 4, 図 5 ともに混合を行っていない $w = 0.0$ (人気エリア法) や $w = 1.0$ (協調フィルタリング) より混合を行った結果が、いずれのデータセットにおいても AUC 値が高く性能が良い結果となっている。さらに、重み w の変化における性能検証の結果、いずれのデータセットにおいても混合するとさらに性能が

上がり、重みが $w = 0.9$ の時に AUC の値がほとんどのデータセットにおいて最大になることがわかる。ところで、 $w = 0.9$ とした混合法の precision 値が図 3 の赤線であり、どのデータセットにおいても他手法より高い値を示しているため、混合法を用いた推薦の有効性が示唆されていると思われる。

図 3(d), (e), (f) において人気エリア法に比べて協調フィルタリングが低い値を示すことについては、ユーザが過去に京都観光しているという京都観光でのリピート性が高い事が嗜好に関係し、予測精度に影響を及ぼしたという仮説が考えられる。また、学習データにおける京都での撮影結果を削除した影響により、学習データ不足によって推薦する観光エリア A_k のスコア AS_k への評価が減少した可能性も考えられる。これらは、今後、対象都市を変えたより詳しい検証が必要であると考える。

5 可視化システムの評価実験

5.1 実験設定

混合法における w について、観光エリア推薦を行った結果から、 $w = 0.9$ が最も推薦結果を高くすることが示された。人気エリア法よりも、嗜好を強く反映させた推薦を行うことが効果的であることを示唆していると思われる。そこで、提案可視化システムにおいて、人気エリア法による可視化結果と、混合法による可視化結果を比較検証するため、一例として、Dataset1 を対象とし、人気エリア法による可視化結果と $w = 0.9$ とした混合法による可視化結果を比較する。

5.2 実験結果

人気エリア法による可視化結果を図 6 に示す。図 6 より、京都駅が最も赤く示されている。しかし、それ以下の観光エリアはあまり目立っていないことがわかる。これは、多くの人々が共通して京都駅を撮影する傾向があるためであると考えられる。一方、提案する混合法の可視化結果を図 7 と図 8 に示す。図 7 では複数の個所が赤く表示されていることがわかる。その観光エリアは図 7 において、(a) 金閣寺、(b) 二条城、(c) 銀閣寺、(d) 清水寺、であった。これらは京都における有名な寺社仏閣であり、伝統的な施設を好む嗜好が可視化に反映されていると思われる。次に図 8 では特定の有名な施設を含む観光エリアではなく、図 8 において、(a) 三条通や (c) 四条通といった店や、古来の街並みを残した通り、(d) 錦市場といった商店街が推薦されており、街歩きや食べ歩きを好む多重嗜好が可視化に反映されていると思われる。以上から、混合法は、ユーザ

の多重嗜好に応じた推薦を行える可能性がある点で提案法の有効性が示唆される。

6 まとめ

大量のメタ情報付写真群を用いて実行動に基づく観光エリアの推薦システムを構築した。提案可視化システムの評価実験により、個人化推薦の観点で有効性を示した。今後は、より多くの都市を対象に評価実験を行い、より洗練された個人化推薦を構築するための探究を行う予定である。

参考文献

- [1] 神島敏弘, “推薦システムのアルゴリズム (2),” 人工知能学会誌, vol.23, no.1, pp.89–103, 2008.
- [2] D.J. Crandall, L. Backstrom, D. Huttenlocher, and J. Kleinberg, “Mapping the world’s photos,” Proceedings of the 18th International Conference on World Wide Web, pp.761–770, 2009.
- [3] S. Kisilevich, F. Mansmann, and D.A. Keim, “P-DBSCAN: A density based clustering algorithm for exploration and analysis of attractive areas using collections of geo-tagged photos,” 1st International Conference on Computing for Geospatial Research & Application, pp.38:1–38:4, 2010.
- [4] S. Kisilevich, F. Mansmann, P. Bak, D.A. Keim, and A. Tchaikin, “Where Would You Go on Your Next Vacation? - A Framework for Visual Exploration of Attractive Places,” GeoProcessing 2010, pp.21–26, Feb. 2010.
- [5] 王 佳な, 野田雅文, 高橋友和, 出口大輔, 井手一郎, 村瀬 洋, “Web 上の大量の写真に対する画像分類による観光マップの作成,” 情報処理学会論文誌, vol.52, no.12, pp.3588–3592, 2011.
- [6] 熊野雅仁, 小関基徳, 小野景子, 木村昌弘, “地理および時間情報をもつ写真データに基づいたホット撮影スポットの抽出,” 情報処理学会論文誌, vol.5, no.3, pp.41–53, Sept. 2012.
- [7] 熊野雅仁, 岩淵聡, 小関基徳, 小野景子, 木村昌弘, “集合知に基づいたポピュラー撮影スポットに関する旬シーズンの可視化,” 芸術科学会論文誌, vol.13, no.4, pp.218–228, 2014.
- [8] D. Comaniciu and P. Meer, “Mean shift: a robust approach toward feature space analysis,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.24, no.5, pp.603–619, 2002.

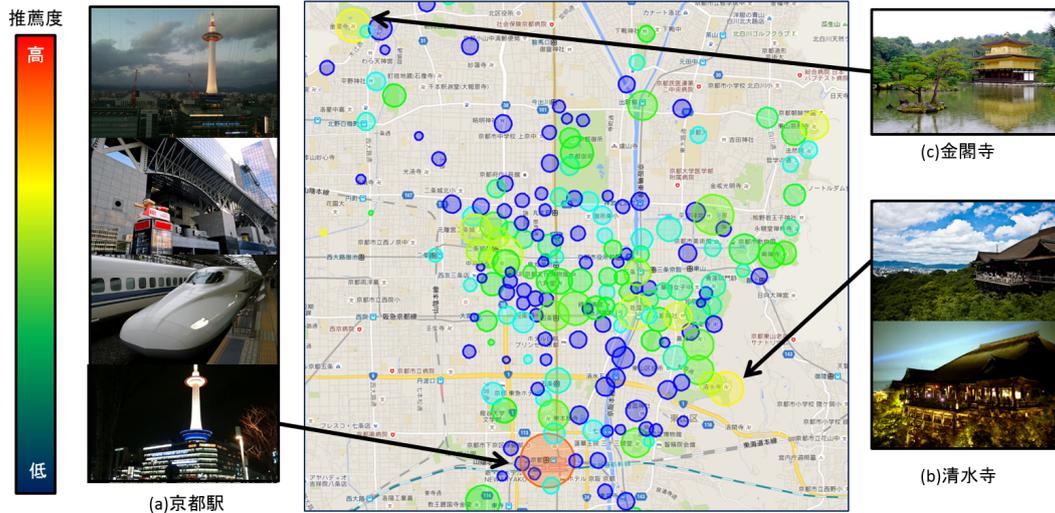


図 6: 人気エリア法を用いた可視化結果

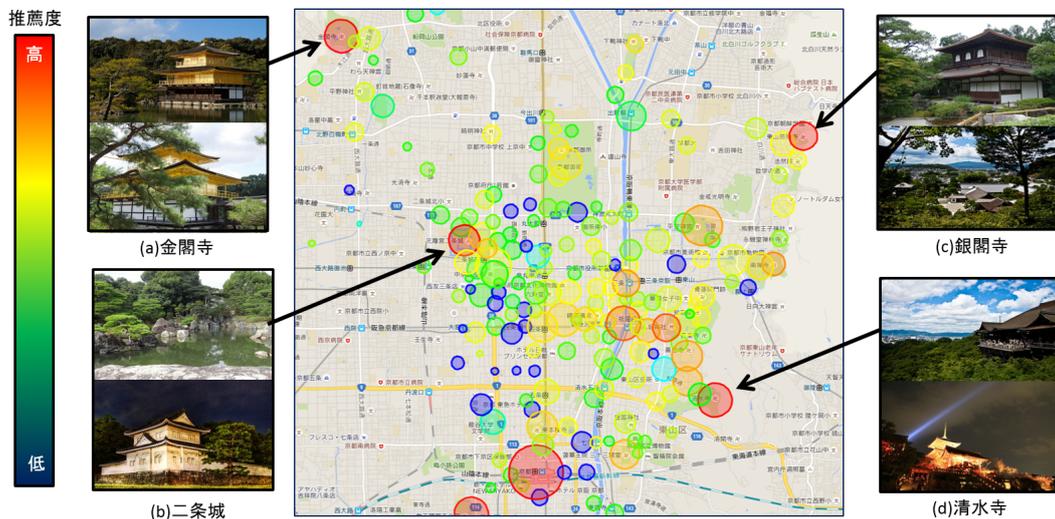


図 7: 提案法を用いた可視化結果 1

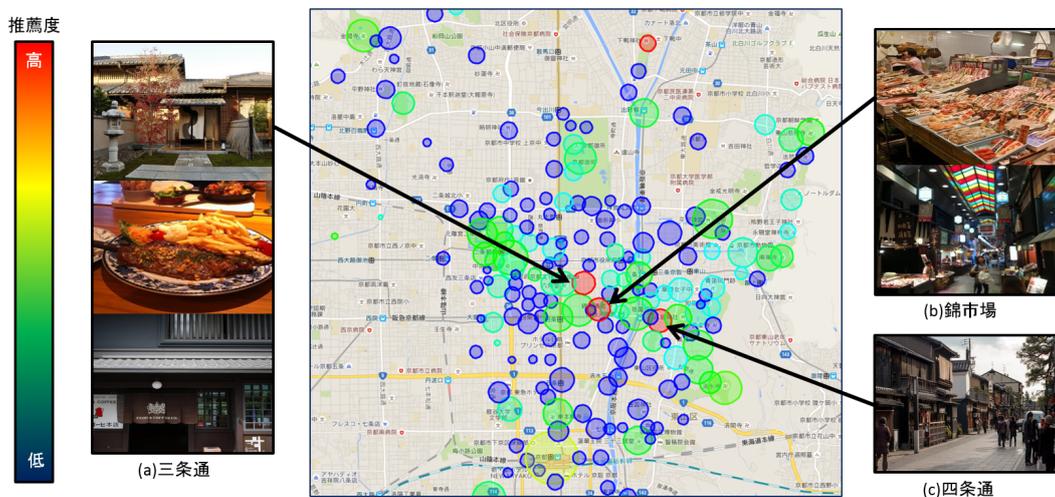


図 8: 提案法を用いた可視化結果 2